DOCUMENT RESUME

ED 425 488 EA 029 175

AUTHOR	Carnevale, Anthony P.; Kimmel, Ernest W.
TITLE	A National Test: Balancing Policy and Technical Issues.
INSTITUTION	Educational Testing Service, Princeton, NJ.
PUB DATE	1997-00-00
NOTE	20p.
AVAILABLE FROM	Educational Testing Service, Mail Stop 01-C, Princeton, NJ 08541.
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Comparative Education; Elementary Education; Grade 4; Grade 8; *Mathematics Achievement; *National Competency Tests; *Reading Achievement; *Test Construction; *Test Results
IDENTIFIERS	*Clinton (Bill)

ABSTRACT

President Clinton, like President Bush before him, has challenged the nation's schools to participate in a rigorous national test of each student's reading skills at grade 4 and mathematics skills at grade 8. Tests would be voluntary and administered by private testing companies. They would allow parents to judge the adequacy of their child's school and to compare their child's performance to that of every other child in the country and in the world. Currently, only 40 percent of American children meet basic reading standards in the fourth grade, and only 20 percent have studied algebra by the eighth grade, compared to 100 percent in many other countries. Proponents see national tests as critical levers to raise the quality of American schooling. Critics contend that testing should be associated with an affirmative strategy to provide all students with learning opportunities. The idea of national tests challenges traditions of local control, the states' role, and use of existing standardized testing programs. Plus linkages to the National Assessment of Educational Progress (NAEP) could be problematic. Also, the greater the consequences, the greater the pressure on test validity, security, and technical quality. High-stakes accountability actions and decisions should reflect multiple kinds of evidence. Test design challenges include adaptation, reliability, performance benchmarks, scoring, reporting results, disclosure, and security. (Contains 32 endnotes.) (MLH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 425 488

A National Test:

Balancing Policy and Technical Issues

Anthony P. Carnevale
Ernest W. Kimmel

EA 029 175

Educational Testing Service

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
P. Carnevale
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

A National Test:

Balancing Policy and Technical Issues

Anthony P. Carnevale
Ernest W. Kimmel

EDUCATIONAL TESTING SERVICE

Mail Stop 01-C
PRINCETON, NJ 08541
Voice (609) 734-5531
Fax (609) 734-1140

1800 K Street, NW, Suite 900
WASHINGTON, DC 20006
Voice (202) 659-8056
Fax (202) 887-0875

Educational Testing Service (ETS) is a private, nonprofit corporation devoted to measurement and research, primarily in the field of education.

Copyright © 1997 by Educational Testing Service, Princeton, NJ

Foreword

President Clinton, like his predecessor President Bush, has argued for common, nationwide tests in reading and mathematics as tools in the struggle to help all students acquire the skills they will need to function in the emerging information-based economy of the United States. No citizen can afford to be without these fundamental skills, nor can American society afford to have its children progress through school without acquiring proficiency in reading and mathematics.

Because Educational Testing Service has been testing students and adults for half a century, a variety of people, ranging from policymakers to parents, have asked us for information that will help them in considering national testing. In response to such queries and drawing on ETS's wide experience in developing and implementing testing programs, two senior members of the ETS Office of Public Leadership have prepared this primer. Anthony Carnevale and Ernie Kimmel have taken special care to capture a wide variety of political viewpoints, as well as the technical issues that need to be considered in any informed debate about the expectations that we set for youngsters in America and the ways we ensure their achievement. The authors have done their best to reflect the full range of views in the current public dialogue, sometimes by letting leaders speak for themselves and sometimes by summarizing arguments made by others.

I commend this essay to your reading. I trust it will be helpful to all concerned about improving the quality of learning in American schools. I know the authors would welcome reactions; therefore, a self-addressed, postage-paid reply form has been included for your convenience in responding.

Nancy Cole
President
Educational Testing Service

What Has the President Proposed?

President Clinton, as President Bush before him, has challenged all of the nation's schools to participate in a rigorous national test of each student's reading skills at grade 4 and mathematics skills at grade 8. The President argues:

What we need are tests that will measure the performance of each and every student, each and every school, each and every district, so that parents and teachers will know how every child is doing compared to other students in other schools, in other states, in other countries — not just compared to them, but more importantly, compared against what they need to know.[1]

Currently, only 40 percent of American children meet basic reading standards in the fourth grade, and only 20 percent have studied algebra by the eighth grade, compared to 100 percent in many other countries.

The proposed common tests would be voluntary and administered by private testing companies. They would allow parents to judge the adequacy of the education provided by their child's school and to compare the performance of their child to that of every other child in the country and in the world. The proposed tests would be limited to reading and mathematics, where there is broadest agreement about what children should know and be able to do. Children need to be proficient in reading in the early grades because it is a tool to learn other subjects. Students need to be proficient in basic mathematics before high school if they are to take rigorous math courses in high school that will prepare them for college and the workplace. Currently, only 40 percent of American children meet basic reading standards in the fourth grade, and only 20 percent have studied algebra by the eighth grade, compared to 100 percent in many other countries.

While polls show strong public support for tougher academic standards and while a far more ambitious testing plan was proposed six years ago by the Bush Administration, some critics have voiced reservations about the President's challenge. As was the case with President Bush's national test proposal, the current proposal has triggered a healthy dialogue on the federal role in education. Rep. Peter Hoekstra's (R-MI) comment following the State of the Union address is indicative: "... before we begin any major new initiatives in the education area, let us take a look at this broad range of Federal programs and find out what is really working and what is not working."[2]

The current dialogue on the merits of a national test reflects the tension between our growing need for ensuring educational performance across the nation and our tradition of local control of education. Proponents of the national test argue that the way forward beyond the current impasse is to educate locally but to assess some common outcomes nationally. By empowering students, their parents, and teachers with information that allows them to assess students' progress

relative to children next door, in the next school district, nationwide, or in other parts of the world, we can strengthen our local schools while encouraging a consistent quality of education and equality of opportunity. In joining the President before the Michigan legislature to endorse the national test proposal, Gov. John Engler argued that, "Now with your initiative, our citizens can know how Michigan children are doing, not just compared to other states, but compared to the world."[3] Ultimately, proponents argue, a national test should empower families by eliminating safe havens from accountability. A well-developed assessment system that is not tied to where learning occurs opens up possibilities for greater diversity in the locus of schooling. Multiple approaches can be used to help children attain a set of common expectations.

But as Congress focuses more directly on the testing proposal itself, it will confront numerous technical issues that must be handled thoughtfully. This paper examines a series of points, both substantive and procedural, that must be addressed in order to ensure that national tests serve as a constructive step toward educational improvement. We offer this primer in the hope that it will inform the current dialogue.

What Purpose Would be Served by a National Test?

By empowering students, their parents, and teachers with information, we can strengthen our local schools while encouraging a consistent quality of education and equality of opportunity.

National tests are seen by their advocates as critical levers to raise the quality of the American educational system to make the U.S. more competitive in the context of a global economy. It is argued that the tests will empower students, parents, teachers, and communities by providing information about the knowledge and skills that are critical for all citizens. In recent years, it has become commonplace to note that global competition begins in the classroom and that our classrooms, like our companies, need to compete globally. Proponents argue that without information that is comparable across the nation and the globe, it is impossible for parents and communities to judge whether their children are being prepared to compete successfully in the global economy. "We can't have excellence without standards, and we can't meet standards without tests," says Frank Doyle, former executive vice president and CEO member at General Electric and current chair of the Committee for Economic Development.[4]

The reading and mathematical skills of American students already are tested repeatedly, primarily with a plethora of tests that do not allow us to compare the performance of individual students across school districts, states, and nations. Our only national and international assessments are periodic low-stakes tests that do not measure individual performance;

thus, they are without serious consequences to students, teachers, administrators, or public officials. This inability to report individual performance in relation to national or international standards hinders accountability. A recent report on the condition of public education in the United States stated: If the data we depend on to monitor the economy were "as incomplete, as unreliable, and as out of date" as the data we depend on to monitor education in the United States, we might as well have the economy of a developing country.[5]

There would be much less concern about having common data if U.S. students were meeting national and global standards, but they are not. Almost half of all our country's 17-year olds fall short of the reading and math skills needed to get a job in a modern automobile plant.[6]

In addition, policymakers argue that they have difficulty in allocating resources because there is no comparative basis for judging the effectiveness of various policies or practices, i.e., to determine where the use of resources is most likely to produce gains in student learning. There are large variations in methods and spending in the nation and in the world. Absent more articulate and recurrent data, it will be difficult to develop actionable information to increase accountability and understand best practices. We could learn a lot about what works if we better understood why the Czech Republic spends a third as much per pupil as we do and ranks 6th in math and 2nd in science while we rank 28th in math and 17th in science.[7]

It is critical that testing be associated with an affirmative strategy to provide all students with an equal opportunity to learn.

In spite of the apparent need for reliable and actionable national assessments, there are good reasons to be cautious and deliberate in constructing a national test. It is critical that testing be associated with an affirmative strategy to provide all students with an equal opportunity to learn. We need to support an educational system offering a route to success that does not depend on one's social standing at birth. Although far from perfect, our public educational system is the best way we have learned, to date, to avoid the constant reproduction of economic and cultural elites.

The federal government has a critical supporting role to play in ensuring consistent quality of education and equal educational opportunity in a mobile society. Within one year, 17 percent of U.S. residents will reside in a different location than they did in the previous year. Mobility among the young tends to be even greater. Thirty-nine percent of second graders have changed schools in the prior two years, and 12 percent changed schools three or more times.

The growing importance of education as the key to a productive and rewarding role in society argues for a means of

monitoring individual accomplishment that will be informative to students, their parents, and teachers. At the same time, however, there will be tremendous pressure for a substantial majority of students to "pass" the national tests. *The Philadelphia Inquirer* pointed out that "Extolling higher expectations and tougher testing is one thing. It is quite another to stomach the reality: that all but the top students tend, at least at first, to come out looking pretty bad."[8] Yet, if low standards on the tests are used to paper over real differences in achievement, and, more importantly, in the opportunity to learn, we will be distracted from finding real solutions to our educational and social problems. Yielding to the temptation to downplay differences in skills will make everybody, including parents, comfortable when they shouldn't be. Further, if we break the critical link between standards and opportunity, we serve neither our interest in global competition nor our interest in equal opportunity.

The content of any national test will spark debate — America's diversity guarantees that. The tests will be viewed in the context of the cultural wars and the demands for sensitivity to differences in the heritage of the diverse groups that comprise our society. There will be arguments about pedagogy, for example, "whole language" versus "phonics" as a strategy for teaching reading.[9] The developers of the proposed tests will need to exercise great care and obtain extensive reviews of the test material in order to minimize the opportunity for such criticisms. Most of the major test publishers have established and proven processes for screening test questions for such cultural, religious, racial/ethnic, or gender sensitivities — and these must be used extensively in developing the national tests.

The content of any national test will spark debate — America's diversity guarantees that.

National Tests and Local Control

To many observers, the very idea of a national test directly challenges the American tradition that education is controlled by states and local school districts — a sentiment captured in remarks by House Appropriations Chairman Robert L. Livingston (R-LA): "The federal school board is not what we need to be."[10] Can a plan for national testing accommodate both our desire for local control and our need for reliable data about student performance against common high standards?

A broader concern is the influence of national tests in shaping state and local education policy. Some, like Lynne Cheney of the American Enterprise Institute, worry that the Clinton reading and math tests are ". . . merely the first step on a path toward central control of all aspects of education."[11]

But as two former education officials of the Reagan and Bush administrations have written: "To those worried about 'local control,' we say that these tests are a yardstick, not a harness. They give the federal government no new powers. The test results, in fact, will actually enhance local control by empowering consumers, policy makers and professionals to know what actions need to be taken locally to improve education."[12]

Currently, we have a hodgepodge of different tests, a hodgepodge of different standards around the country.

Indeed, for many advocates of national tests, it is parents — the most local stakeholders of all — who have the most to gain. "[W]e need to give parents clear indication of which schools are doing the best job in educating students. Currently, we have a hodgepodge of different tests, a hodgepodge of different standards around the country. Parents who are interested in finding out how their children are doing often are misled by inaccurate information," said Sen. Jeff Bingaman (D-NM), in response to the President's State of the Union address.[13]

Rep. Frank Riggs (R-CA), chairman of the House Subcommittee on Early Childhood, Youth and Families, may have best expressed the cautious attitude of many observers in saying that testing is ". . . one of the more intriguing parts of [the President's] education proposal." While raising concerns about the Administration's plan to proceed without explicit congressional authorization, he acknowledged: "We need some sort of standardized assessment and performance-based assessment to ensure learning in the core academic subjects."[14]

The State Role. Among state-level officials, the response has been similar. The Council of Chief State School Officers has welcomed the plan, but individual superintendents say three areas need to be addressed: "avoiding duplication with existing state tests; ensuring the reliability of the tests during the fast-track development schedule envisioned by the administration; and keeping costs down."[15] Among 27 state agencies contacted by *Education Daily*, a wait-and-see attitude prevails.[16] Some officials welcome the opportunity to benchmark their own programs against national exams (or possibly, to substitute the national tests for their own). But others worry that any national test will be calibrated to the lowest common denominator. Delaware's assessment director, Rebecca Kopriva, says that once a national test is available, ". . . we can't deviate from it, even if we think our test is better."[17]

Variability in state progress on standards and assessments is one reason for a nationally calibrated test. According to the National Education Goals Panel, ". . . what is considered 'good enough' for student performance varies from state to state."[18] The Administration contends that "while most states assess their students in reading and math, they generally set their

Variability in state progress on standards and assessments is one reason for a nationally calibrated test.

proficiency levels lower than the challenging levels in the NAEP [National Assessment of Educational Progress]."[19] This view is echoed by Rep. Harris W. Fawell (R-IL), who agrees that "states should be the primary source of quality education" but adds that "over the years the states have . . . pretty much ignored rigorous course standards in testing" and that they need "incentives" to improve.[20]

Several Tests, One Standard? Since the reading and math skills of American students already are tested repeatedly, with the vast majority of students taking tests produced by three major commercial publishers, some commentators have suggested that the federal government should collaborate with the testing industry to develop a means of linking existing tests to NAEP proficiency levels. This would preserve the aura of local control by permitting each state or district to continue using its preferred test; it could also preserve competition among commercial test publishers without a government-sponsored test appearing to challenge their legitimacy.

Unfortunately, linking disparate tests to each other and to NAEP would almost certainly fail to provide dependable results. Robert Mislevy, an ETS researcher who has worked extensively on test linkage,[21] calls this idea "educational assessment's counterpart to perpetual motion machines. It is an appealing and compelling idea, eminently plausible on the surface. It unavoidably leads to disappointment, at best, or disaster if there are stakes attached to the results."[22]

To explain Mislevy's misgivings, consider some of the knots that will arise in attempting to correlate commercial tests with NAEP scales:

- Schools often use commercial tests for several years, and may inflate their scores by "teaching to the test." Transforming such results to a common proficiency scale would do nothing to make them more trustworthy. (In fact, correcting this problem is a primary reason for standards-based national testing.)
- Existing tests proposed for linkage to NAEP rely heavily on multiple-choice questions, while NAEP consists largely of constructed-response items, i.e., problem-solving or essay questions.
- Finally, the commercially available tests are not only different from NAEP but also from each other, making linkage even more difficult.

Attractive as this option sounds, it would probably lead to incorrect inferences and conclusions about the actual state of student learning — and defeat the very purpose of the initiative.

Nationwide Tests and National Standards

For some, the problem is not national tests as such; it is that any national testing system necessarily depends on common national expectations about reading and math performance at the grade levels tested. Will national standards shoehorn our diverse nation into a "one size fits all" approach?

On one side of this issue are those such as Education Secretary Richard W. Riley, who hold that "reading is reading and math is math, whether we are in Maine, Missouri, or Montana."[23] On the other side are those who recall the stormy reception given some previous national standards efforts and warn, in the words of an aide to California Gov. Pete Wilson: "We don't think you can set national standards that can be considered world-class. . . . When you factor in all the concerns of all the states, you end up with a watered-down product."[24]

What is not being proposed are nationally mandated standards but ... national standards that can be adopted locally and carried out locally.

To Rep. John E. Porter (R-IL), the standard-setting required by national testing should be understood in a different light: "It seems to me we need . . . a shorthand way of explaining . . . that what is not being proposed are nationally mandated standards but . . . national standards that can be adopted locally and carried out locally." He suggests using the phrase "state-adopted national standards."[25]

The Administration makes a distinction between "federal government standards . . . imposed or required by the federal government," and "national, voluntary standards developed by groups of individuals outside of government." It cites, in particular, the National Assessment Governing Board (NAGB), whose widely respected NAEP frameworks were developed with wide public and professional input and reflect broad national consensus about how well students should read in fourth grade, or do math by eighth. It is these frameworks that will serve as the basis for the national tests.

To Chester E. Finn, Jr., an assistant secretary of education during the Reagan Administration, the NAEP connection helps resolve the "national standard" problem. In testimony on the testing proposal, he noted: "The Clinton plan does not envision brand-new tests or standards. It starts with the NAEP tests that we already have and the NAGB standards already built into them. . . . All that the President is basically proposing is to take two well-regarded existing tests of core skills, tests that already contain generally-accepted national standards, and allow these to be used more widely by those who wish to do so."[26]

National expectations, local options. When Secretary Riley says "reading is reading," some say that's not so — that the

field is rife with controversy about whether phonics or whole-language is the right way to teach that essential subject. True enough — but the practical effect of national testing will be to measure results, not dictate how to achieve them. However, it is important to recognize that those who build the test must have a clear idea of the expectations for students at the educational level being tested. The tests necessarily will be based on either implicit or explicit definitions of what students should know or be able to do — a *de facto* national standard.

While the recent history of national standard-setting is fraught with difficulty, the Administration has avoided at least some controversy by limiting its proposal to testing in two basic subjects. Moreover, it has chosen grade levels at which student achievement in these areas is regarded as an important "gatekeeper." According to Christopher T. Cross, president of the Council for Basic Education: "Research shows that students unable to read well by the end of the 3rd grade are more likely to become dropouts, struggle in later grades, and have fewer good job options. Developing math proficiency in junior high school enables students to succeed in rigorous math and science courses in high school, which have been shown to increase scores on college-entrance exams and prepare students for the intellectual challenges of college and careers."[27]

We should recognize that the potential uses must be a major determinant of the test design. The greater the consequences of the test results — for student, teacher, or school — the greater the pressure on its validity, security, and technical quality.

This emphasis on essential academic skills may be one reason the testing proposal has received an enthusiastic reception from the business community. Not only did 200 executives, many from high-technology firms, join with the President in a White House event to promote the testing idea, but major national organizations such as The Business Roundtable and the U.S. Chamber of Commerce have also embraced it. Some have suggested that the business community should play an active role in test development, perhaps providing real-world workplace problems that can be solved using eighth-grade math skills.

How Will the Tests be Used?

Rewards or sanctions — based on test results — can have a powerful effect on the way in which a new testing program is implemented. If all a national test would do is report on a student's learning progress, pressures to distort the process would be minimal. However, we are likely to use national test information for many other purposes as well. We should be clear about which purposes national tests might usefully serve and recognize that the potential uses must be a major determinant of the test design. The greater the consequences of the test results — for student, teacher, or school — the greater the pressure on its validity, security, and technical quality.

Competition and Consequences. The President's recent remarks describe one reason for a national measure of student accomplishment — to make comparisons among students in different schools or states against an external standard of what should be learned. In the President's view, this should stimulate healthy competition and stronger results.

Yet in a recent hearing to consider the President's education proposals, Rep. William F. Goodling (R-PA) Chairman of the House Committee on Education and the Workforce — himself an educator — expressed misgivings about test-based comparisons: "My whole idea as an administrator was to have teachers use tests to determine where they didn't get the point across to the student. . . . But obviously, the whole world will know the results [under this proposal] and then you'll start measuring one school district in relationship to another school district. And I have real problems with that. . . ."[28]

Return to Lake Wobegon? A second issue is raised by the tests' voluntary nature. NAEP results are expressed as "proficiency levels" designed for reporting broad trends based on a carefully chosen sample of students. But the new tests will be designed to report achievement for individual students, then aggregated up to classroom, school, and district levels. If only the best-prepared schools choose to take part, the results could paint an inaccurate picture that most districts are "above average" — just like the children in Garrison Keillor's mythic hometown.

Opportunity to Learn. Equity is a third concern. As Rep. Louis Stokes (D-OH) recently asked, recalling visits to two schools with widely varying per-pupil spending: "How can we do national testing with this type of disparity in the school systems in this country?"[29] If any jurisdiction bases significant decisions about individuals on the tests' results, it may expect scrutiny. Districts may be taken to court to defend the technical quality of the test and to demonstrate that students had the opportunity to learn the skills it measures.

High-stakes accountability actions, and decisions that have a significant impact on individual students, should reflect multiple kinds of evidence.

For individual students, the tests may mean high stakes indeed. As currently envisioned by some, districts could use test results to decide whether a student is promoted to fifth grade or retained in the fourth. For teachers, principals, and entire school districts, aggregated test results might mean rewards for performance and sanctions if their students fall short. Each of these uses poses additional challenges for test security, technical justification, and maintenance of integrity. And they call for some caution. Any one test can provide only a snapshot of a student's performance on a limited sample of subject matter. High-stakes accountability actions, and decisions that have a significant impact on individual students, should reflect multiple kinds of evidence.

Design Challenges

How best to address these thorny issues while moving ahead with dispatch? The President's plan will create expert advisory groups to consult on design questions. Among the challenges they will face are the following:

Adaptation. NAEP content frameworks were designed for tests that use a complex sampling design in which no single student is tested on all aspects of the framework. The new proposal calls for each student to cover the whole domain (or a representative sample of it) within a 90-minute time limit. Thus, the specifications cannot be identical to NAEP; they will require rethinking the framework. Will the prescribed format and time constraints lead to tests that are so different from NAEP that the desired links to the proficiency scales cannot be made?

Reliability. If the tests consist of 80 percent multiple-choice and 20 percent constructed-response (problem-solving or essay) questions, will they be reliable enough for the anticipated individual uses? This will be a particular issue in reading if the longer passages characteristic of NAEP are used and students are asked to complete work within the proposed 90-minute time frame.

Performance Benchmarks. Early reports said the Administration planned to use NAEP's "basic" level as a national standard for student performance, rather than the "proficient" level of NAEP's higher achievers. The Administration has clarified its position: "We intend to ensure that students' scores can be reported in a manner that permits parents and teachers to know whether the students have attained NAGB's 'basic,' 'proficient,' or 'advanced' levels. We believe, as does NAGB, that all children should be at least 'proficient' in basics and other subjects."[30] The advisory groups must make clear what constitutes acceptable student performance at each grade level on each of the tests.

Scoring. How can student performance on the open-ended items be evaluated consistently when, as proposed, the scoring is done by various local or state groups? If constructed-response questions hinge on only a few key words, it may be relatively easy to score them consistently across different schools and districts. But if the items are such that several different kinds of answers might be considered satisfactory, as is the case with most complex performances, scorers will require extensive training to ensure comparability in their judgments.

Reporting the Results. Should the tests be crafted to ensure reliability throughout the reporting scale or to maximize the

reliability of classifying students as "basic," "proficient," or "advanced"? This decision will affect, for example, how many questions of different levels of difficulty are used.

Validation. Ensuring that the tests are valid measures of fourth-grade reading and eighth-grade mathematics will require many steps. Some of the most prominent considerations include:

- reviewing the content in light of the emerging standards;
- ensuring the clarity and comparability of the test administration procedures at all schools;
- evaluating the appropriateness of test conditions to age, disabling condition, or English-language proficiency; and
- relating the new tests to other measures of reading and math accomplishment.

In addition, the implications of the tests being administered on a voluntary basis must be judged for each of the validity considerations.

Students (and their teachers) have a right to see a set of disclosed questions prior to taking the test, and test takers should have no previous exposure to the test questions.

Disclosure and Security. The Administration has stated that the test will be disclosed after it is administered annually, "so students, parents, and teachers can know what is necessary to reach standards of excellence."[31] This is an important goal — but it must be implemented within the broader context of who has access to the test questions and when. The issues of disclosure and test security are intertwined; few would dispute the twin principles that:

- students (and their teachers) have a right to see a set of disclosed questions prior to taking the test, and
- test takers should have no previous exposure to the test questions.

However, implementing the first principle, while ensuring the second, is not a simple matter.

According to the President's proposal: "The tests will be licensed to states, school districts, and test publishers. These licensees will be responsible for administering and scoring the tests to ensure that standards of validity and reliability are met."[32] Can the security of the tests be adequately maintained with this voluntary, decentralized approach? A system of multiple licensees would entail many different streams of production, printing, and shipping, leaving many opportunities for premature disclosure of test items. And if the tests are to be used for significant decisions about individuals, there

will be strong pressures on students (and teachers) to obtain prior knowledge of the questions — a temptation enhanced by electronic communications.

If multiple forms of the test are given each year, there will need to be secure questions, common to more than one form, that can be used to equate those forms, that is, to ensure the comparability of scores received on any of the forms. Similarly, to ensure the comparability of scores from one year to the next, it will be important to have secure questions that can be used in more than one year. This technical need to have secure questions that can be administered on more than one occasion argues for the disclosure of only a portion of each test following its administration.

If any one of these technical issues is mishandled, public confidence in the results may erode, and this could impair the perceived legitimacy of the national tests or even of NAEP data. The advisory groups must conduct their deliberations impeccably and include members with superb technical skills.

Tests and Learning: Acting on Results

If national testing is done right, it will create unprecedented demand to act on what we discover. Measuring a child's height does not help him or her to grow taller; it only tells where that child is in relation to others of the same age group, and may say something about genes, nutrition, and overall health. Similarly, just measuring reading and math proficiency does little to develop those skills; it just tells where a child is in relation to grade-level expectations and reflects the learning opportunities the child has had.

To improve learning, test results must be put to use!

To improve learning, test results must be put to use! They need to be in a form that parents can understand; they should provide teachers with useful information; and they should help schools and school districts to judge, over time, the effectiveness of their decisions about curricula, technology, and textbooks. To provide these benefits, the supplementary materials that accompany the national test results will be as important as the tests themselves.

For parents, educators, and policymakers, the potential value of new national tests lies in knowing whether students are truly achieving, and where there are gaps in their learning — especially in the vital basics of reading and math — and, eventually, how to close those gaps.

The technical issues are difficult, but surely surmountable.

Notes

1. Transcript of President Clinton's remarks to the Maryland State Legislature, February 10, 1997.
2. Hon. Peter Hoekstra, remarks on House floor, February 5, 1997 (*Congressional Record*, p. H335).
3. Hon. John Engler, "A Passion for Excellence in Education." Address on the occasion of President Clinton's visit to a joint session of the Michigan legislature, March 6, 1997.
4. Frank Doyle, personal communication.
5. *Quality Counts, Education Week* special supplement, January 22, 1997, p. 18.
6. R. J. Murnane and F. Levy, *Teaching the New Basic Skills*, New York, NY: The Free Press, 1996, p. 35.
7. U. S. Department of Education, National Center for Education Statistics, *Pursuing Excellence*, NCES 97-198, Washington, DC: U. S. Government Printing Office, 1996, pp. 20-21.
8. Dale Mezzacappa, "Behind the Delay: Why the Schools' Standards are Still Low," *The Philadelphia Inquirer*, March 30, 1997, p. 1.
9. Lynne V. Cheney, "Whose National Standards?," *The Wall Street Journal*, April 2, 1997.
10. David J. Hoff, "Clinton Gives Top Billing to Education Plan," *Education Week*, February 12, 1997.
11. Lynne V. Cheney, "Whose National Standards?," *The Wall Street Journal*, April 2, 1997.
12. Chester E. Finn and Diane Ravitch, "A Yardstick for American Students," *The Washington Post*, February 25, 1997.
13. Sen. Jeff Bingaman, remarks on Senate floor, February 5, 1997 (*Congressional Record* p. S991).
14. David J. Hoff, "Political Shift Emboldens Clinton to Urge Tests," *Education Week*, February 19, 1997.
15. David J. Hoff, "Chiefs' Group Backs Clinton Testing Proposals," *Education Week*, March 26, 1997.
16. Laureen Lazarovici, "States Still Ambivalent About National Tests," *Education Daily*, Vol. 30, No. 68, April 9, 1997, p. 1.
17. Quoted in "States Fear National Tests Will Thwart Their Reforms," *What Works in Teaching and Learning*, Vol. 29, No. 6, March 12, 1997, p. 5.
18. National Education Goals Panel, The National Education Goals Report — *Building A Nation of Learners*, Washington, DC, 1996, p. 8.

19. "Mastering the Basics Once and for All: Reaching for High Standards in Reading and Math." Press materials on the President's Initiative on National Standards of Excellence, February 1997.

20. Hon. Harris W. Fawell, remarks, House Committee on Education and the Workforce hearing, March 13, 1997.

21. R. J. Mislevy, *Linking Educational Assessments: Concepts, Issues, Methods, and Prospects,* Princeton, NJ: Policy Information Center, Educational Testing Service, December 1992.

22. R. J. Mislevy, personal communication.

23. Secretary Richard W. Riley, testimony before House Committee on Appropriations, Subcommittee on Labor, Health and Human Services, Education and Related Agencies hearing, March 19, 1997.

24. Dan Edwards, California Office of Child Development and Education, quoted in David J. Hoff, "California Officials Divided on Clinton Test Plan," *Education Week,* April 9, 1997.

25. Hon. John E. Porter, remarks, House Committee on Appropriations, Subcommittee on Labor, Health and Human Services, Education and Related Agencies, March 19, 1997.

26. Chester E. Finn, Jr., "Appraising the Clinton Education Plan." Testimony before the House Committee on Education and the Workforce hearing, March 13, 1997.

27. Christopher T. Cross and Scott Joftus, "Stumping for Standards," *Education Week,* April 9, 1997, p. 41.

28. Hon. William F. Goodling, remarks, House Committee on Education and the Workforce hearing, March 13, 1997.

29. Hon. Louis Stokes, remarks, House Committee on Appropriations, Subcommittee on Labor, Health and Human Services, Education and Related Agencies, March 19, 1997.

30. Hon. Marshall S. Smith, Acting Deputy Secretary of Education, letter to The Honorable William F. Goodling, March 19, 1997.

31. "National Standards of Academic Excellence." White paper on President's proposal, via Education Department homepage, April 7, 1997.

32. "Technical Information on the National Reading and Math Tests." Press materials on the President's Initiative on National Standards of Excellence, February 1997.

P/J 266-40

NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

BUSINESS REPLY MAIL
FIRST-CLASS MAIL PERMIT NO. 89 PRINCETON, NJ

POSTAGE WILL BE PAID BY ADDRESSEE

OFFICE OF PUBLIC LEADERSHIP
EDUCATIONAL TESTING SERVICE
MAIL STOP 01-C
PO BOX 6666
PRINCETON NJ 08543-6923

WHAT DO YOU THINK?

With this paper and other efforts of the Educational Testing Service public leadership initiative, we invite you to join in a conversation with us about public educational policy issues that affect you and your community. We would appreciate your thoughts on the issues raised in this paper. Let us know if you would like to receive similar upcoming ETS publications. Please take a few moments to share your views on the proposed national tests and then either fold and put this postage-paid form in the mail or fax it to us at 1-609-734-1140 or 1-202-887-0875. We look forward to hearing from you!

What is your view of the desirability of national tests in 4th Grade English and 8th Grade Mathematics?

What do you consider the most persuasive arguments in support of the proposed tests?

What do you consider the most persuasive arguments against the proposed tests?

Other comments about this issue and/or the way it is treated in this paper.

Would you like to receive future ETS publications about public issues affecting education?

☐ Yes ☐ No

Name: _____

Title: _____

Institution or Organization: _____

Street Address: _____

City/State/ZIP Code: _____

Telephone: (____)_____ E-Mail: _____

Fax: (____)_____

BEST COPY AVAILABLE

28640-14717 • Y67M8.5 • 212001 • Printed in U.S.A.